Lecture 9: October 3rd 2024

Scribe: Adam Wills *Converse to Channel Capacity Theorems and Applications*

9.0 Recap from Last Time

Last time we started the proof of the converse of our channel capacity theorems, both for the classical channel capacity (Shannon's noisy channel coding theorem) and for CQ channels (the HSW theorem). We started to make preparations in this direction by stating Fano's inequality, and we started to discuss the conditional mutual information (CMI). We will continue these discussions now to prove the converse.

9.1 Proof of the Converses

Let us start by considering the general encode - channel - decode scenario:

$$M \overset{\text{Encode}}{\longrightarrow} X^n \overset{\text{Channel}}{\longrightarrow} Y^n \overset{\text{Decode}}{\longrightarrow} \hat{M} \tag{9.1}$$

In a successful protocol, we have that the probability M and $\hat{M}$ differ is at most ϵ. Also, as usual, we say that the alphabet from which M is drawn is of size 2^{nR}. We can make a first step in our proof of a converse by applying Fano's inequality. Considering a uniformly random initial message, M, we have $H(M) = nR$. Then, the mutual information between M and $\hat{M}$ is

$$I(M : \hat{M}) = H(M) - H(M|\hat{M}) \geq (1 - \epsilon)nR - 1, \tag{9.2}$$

where we have applied Fano's inequality. We wish to turn this into a statement about the mutual information between X^n and Y^n, and for this we will talk about the CMI.

9.1.1 Conditional Mutual Information (CMI)

Either quantumly or classically, we can define the CMI as

$$I(X : Y|Z) = H(XZ) + H(YZ) - H(XYZ) - H(Z), \tag{9.3}$$

where quantumly the H's are replaced by S's. Classically we can re-write this definition as

$$I(X : Y|Z) = \sum_z I(X : Y|Z = z)p_Z(z), \tag{9.4}$$

although conditioning in this way doesn't make sense quantumly. It's interesting to notice that this expression means that showing the CMI is non-negative classically is no harder than showing the regular mutual information is non-negative, but because this expression isn't meaningful quantumly, one must work a little harder to show non-negativity of CMI quantumly.

With a little rearrangement of entropies, one can quickly see that

$$I(X : Y|Z) = I(X : YZ) - I(X : Z), \tag{9.5}$$

for which the further rearrangement

$$I(X:YZ) = I(X:Y|Z) + I(X:Z) \tag{9.6}$$

gives us the "chain rule" of mutual information. These expressions lend further weight to the interpretation of the CMI $I(X:Y|Z)$ as the amount of information shared by X and Y once you have conditioned on Z. The equation (9.5) tells us that $I(X:Y|Z)$ is the amount of information X knows about YZ that it doesn't already know about Z. Therefore, $I(X:Y|Z)$ is zero if and only if all of the interactions between X and Y are mediated by Z, which says exactly that $X \to Z \to Y$ forms a Markov chain, as stated last time. Classically, this is easy to prove, whereas quantumly we take it as the definition of a quantum Markov chain.

Let us put some more meat on this Markov chains idea in the classical case. We have

$$I(X:Y|Z) = 0 \iff X \to Z \to Y \text{ is a Markov chain} \tag{9.7}$$

$$\iff p(x,y,z) = p(z)p(x|z)p(y|z) = p(x)p(z|x)p(y|z) \tag{9.8}$$

$$\iff p(x,y,z) = f(x,z)g(y,z) \tag{9.9}$$

for some functions f, g. There is a corresponding robustness statement, which is

$$I(X:Y|Z)_p = \min_{q:q \text{ is a Markov Chain}} D(p||q), \tag{9.10}$$

so that if the joint distribution of X, Y and Z is near a Markov chain (in relative entropy), then the CMI is small. This fits into our general family of similar statements:

$$H(X) \approx 0 \iff X \text{ nearly deterministic} \tag{9.11}$$

$$S(\rho) \approx 0 \iff \rho \text{ nearly pure} \tag{9.12}$$

$$H(X|Y) = 0 \iff X \text{ is a deterministic function of } Y \tag{9.13}$$

$$S(X|Y) = 0 \text{ is an exception — no special meaning!} \tag{9.14}$$

$$D(\rho||\sigma) \approx 0 \iff \rho \text{ is almost } \sigma \tag{9.15}$$

We can also provide a physical interpretation to the CMI being zero. Thinking of the random variables as physical systems interacting, we find that if $I(X:Y|Z) = 0$ then the chain $X - Z - Y$ has only local interactions. Thinking in terms of statistical mechanics, their probability distribution factorises into two Gibbs distributions:

$$p(x,y,z) = \frac{e^{-E_1(x,z)-E_2(y,z)}}{Z}. \tag{9.16}$$

This idea extends to more general networks. Suppose we have physical systems X, Z, Y and W interacting locally via the network shown in the Figure. Removing the system Z, or conditioning

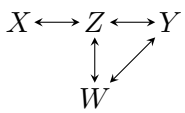


Figure 9.1:

on it, leads to X and W being independent, meaning that the CMI $I(X:W|Z)$ is zero — all interactions between X and W are mediated (in some way) by Z.

Note that not all our discussion of CMI extends to the quantum case. Again, our definition of a quantum Markov chain is simply a state whose CMI is zero, and it is harder to make a robustness statement like Equation (9.10) in the quantum case. Again, non-negativity of CMI is true, but harder to prove. To do so, let us discuss the Data Processing Inequality.

9.1.2 Data Processing Inequality

We will see more on the DPI shortly, but our first form of the DPI will be a classical statement about the Markov chain $X - Z - Y$, for which we have

$$I(X : Z) \geq I(X : Y), \tag{9.17}$$

i.e., the information shared between X and Z is always at least the information shared between X and Y. This is easily proved using the non-negativity of CMI.

Proof.

$$I(X : Z) = I(X : YZ) - I(X : Y|Z) \tag{9.18}$$
$$I(X : Y) = I(X : YZ) - I(X : Z|Y) \tag{9.19}$$

and so

$$I(X : Z) - I(X : Y) = -I(X : Y|Z) + I(X : Z|Y). \tag{9.20}$$

We know the first term is zero using the fact that $X - Z - Y$ forms a Markov chain, and the second term is non-negative, giving the conclusion. □

9.1.3 Strong Subadditivity

Showing that the CMI is non-negative quantumly is harder than classically, and is equivalent to the statement of strong subadditivity, which is the statement that

$$S(XZ) + S(YZ) \geq S(XYZ) + S(Z), \tag{9.21}$$

which is stronger than the usual statement of subadditivity, which recall is

$$S(X) + S(Y) \geq S(XY), \tag{9.22}$$

which itself is equivalent to the regular mutual information being non-negative. The non-negativity of the quantum CMI was initially proved by Lieb and Ruskai in the 70s, which is a different proof to what we show now.

We want to use the fact that, given a quantum operation $\mathcal{E}$,

$$D(\mathcal{E}(\rho)||\mathcal{E}(\sigma)) \leq D(\rho||\sigma). \tag{9.23}$$

This actually follows immediately from our proof of the operational interpretation of $D(\rho||\sigma)$ as the optimal rate at which ρ and σ can be distinguished in an asymmetric hypothesis test. When trying to distinguish ρ and σ, one thing you could do is apply $\mathcal{E}$ to both of them, and then distinguish the resulting states. This immediately gives us this monotonicity statement.

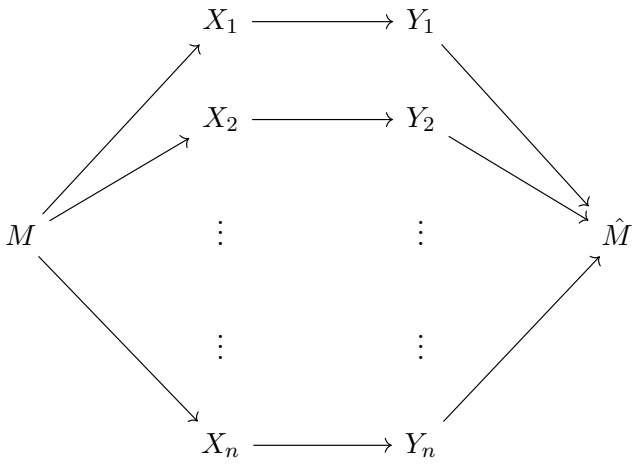
Applying this with $\mathcal{E}$ a partial trace over the system Y, we have

$$I(X : Z) = D(\rho_{XZ}||\rho_X \otimes \rho_Z) \leq D(\rho_{XYZ}||\rho_X \otimes \rho_{YZ}) = I(X : YZ), \tag{9.24}$$

and the result follows.

9.1.4 Back to Channel Capacities

Recall we have the dependence diagram of random variables



which can also be simplified to

$$M \to X^n \to Y^n \to \hat{M}. \tag{9.25}$$

We were up to the point in the proof of the converse where we had $(1-\epsilon)nR - 1 \leq I(M : \hat{M})$, and then using DPI twice gives $I(M : \hat{M}) \leq I(X^n : Y^n)$. We then use

$$I(X^n : Y^n) \leq \sum_{j=1}^{n} I(X_j : Y_j) \tag{9.26}$$

which we will show to be true momentarily. The proof of the converse follows immediately, since we find that we have

$$(1-\epsilon)nR - 1 \leq n \max_p I(X_1 : Y_1)_p, \tag{9.27}$$

so that we find that R is indeed at most the claimed capacity

$$C(\mathcal{N}) = \max_p I(X : Y)_p. \tag{9.28}$$

Let us now prove the claim of Equation (9.26).

Proof.

$$I(X^n : Y^n) = H(Y^n) - H(Y^n|X^n) \tag{9.29}$$

$$\leq \sum_{j=1}^{n} H(Y_j) - H(Y^n|X^n) \tag{9.30}$$

by subadditivity. Then, by a chain rule/telescoping sum,

$$H(Y^n|X^n) = \sum_{j=1}^{n} H(Y_j|X^n Y_1 Y_2 \dots Y_{j-1}). \tag{9.31}$$

Referring to our dependency diagram, we can see that once we condition on X_j, further conditioning via the other $X's$ or any other Y's does not change the distribution of Y_j (this is the definition of such dependence). We therefore have

$$H(Y_j|X^n Y_1 Y_2 \dots Y_{j-1}) = H(Y_j|X_j). \tag{9.32}$$

In total,

$$I(X^n : Y^n) \leq \sum_{j=1}^{n} H(Y_j) - H(Y_j|X_j) = \sum_{j=1}^{n} I(X_j : Y_j). \tag{9.33}$$

□

This concludes our proof of the converse for Shannon's noisy channel coding theorem.

9.1.5 Converse to the HSW Theorem

It turns out that the above argument goes straight through to the case of classical-quantum channels relevant to the HSW theorem, because the inputs are classical. The system Y^n becomes the quantum system Q^n, and we have

$$I(X^n : Q^n) = S(Q^n) - S(Q^n|X^n) \tag{9.34}$$

$$\leq \sum_j S(Q_j) - S(Q_j|S_j) \tag{9.35}$$

$$\leq n\chi, \tag{9.36}$$

where $\chi = \max_p I(X : Q)_\omega$, and we recall the classical-quantum state

$$\omega^{XQ} = \sum_x p(x) \left|x\right\rangle \left\langle x\right|^X \otimes \rho_x^Q. \tag{9.37}$$

This only goes through because the system X is classical - so conditioning on it is sensible. This argument does not go through in the case of entangled inputs, i.e., for a general quantum-quantum channel.

9.2 General Quantum Channels

For a general quantum channel given by the CPTP map $\mathcal{N}$, the capacity is in fact

$$C(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}), \tag{9.38}$$

where

$$\chi(\mathcal{E}) = \max_{\{p(x),\sigma_x\}} I(X : Q)_\omega, \tag{9.39}$$

for $\omega = \sum_x p(x) \left|x\right\rangle \left\langle x\right|^X \otimes \mathcal{N}(\sigma_x)^Q$. The limit in the formula reflects the fact that in general the quantum capacity is super additive, i.e., $\chi(\mathcal{N} \otimes \mathcal{M}) \geq \chi(\mathcal{N}) + \chi(\mathcal{M})$ always, and there are cases where this inequality is strict. This makes the formula quite difficult to work with, a fact also reflected by the optimisation problem contained in it being NP-hard to solve in the worst case. There are, however, several natural cases in which the quantum capacity is additive, meaning that $\chi(\mathcal{N}^{\otimes n}) = n\chi(\mathcal{N})$, and in these cases the formula becomes much easier to work with. Let us state some examples.

1. For entanglement-breaking channels, also called QCQ channels, or measure-and-prepare channels, the capacity is additive. These are channels of the form

$$N(\rho) = \sum_k Tr(\rho M_k)\sigma_k \tag{9.40}$$

for some measurement $\{M_k\}$ and states σ_k. Note that these are more general than CQ channels, for which the first measurement is in the computational basis.

2. Depolarising channel
$$\mathcal{N}(\rho) = (1-p)\rho + p\frac{I}{d}. \tag{9.41}$$
3. Erasure channel
$$\mathcal{N}(\rho) = (1-p)\rho + p\left|e\right\rangle\left\langle e\right|, \tag{9.42}$$
where $|e\rangle$ is some erasure symbol.
4. Unital qubit channels (such channels satisfy $\mathcal{N}(I) = I$ and have one qubit as their input and output).
5. Pure Loss Bosonic Channels. A Bosonic channel can be thought of as acting on the Hilbert space of a harmonic oscillator.

9.3 Random Access Coding

A natural application of these ideas comes in random access coding, specifically the related quantum no-go theorem. The task is as follows. Suppose Alice wishes to encode m bits $x^m \in \{0,1\}^m$ in an n-qubit quantum state ρ_x. She sends this to Bob. Bob wishes to retrieve just one of the bits, of his choosing, say i, by performing a measurement. He learns some bit $\hat{x}_i$ by performing a measurement, and the hope is that $\hat{x}_i = x_i$. One naturally asks to what extent this can be done reliably with various values of n and m.

At finite lengths, you can do a little better with quantum states than you can do classically. For example, clearly, if you encode 2 bits into 1 bit, there is only a 50% probability that Bob can learn a bit of his choosing. However, suppose that Alice encodes 00 into $|0\rangle$, 01 into $|+\rangle$, 10 into $|-\rangle$ and 11 into $|1\rangle$. Then, if Bob wishes to learn the first bit, he needs to distinguish
$$\frac{|0\rangle\langle 0| + |+\rangle\langle +|}{2} \text{ from } \frac{|-\rangle\langle -| + |1\rangle\langle 1|}{2} \tag{9.43}$$
and if he wishes to learn the second bit, he needs to distinguish
$$\frac{|0\rangle\langle 0| + |-\rangle\langle -|}{2} \text{ from } \frac{|+\rangle\langle +| + |1\rangle\langle 1|}{2}, \tag{9.44}$$
and in both cases he can succeed with probability $\cos^2 \pi/8 > 0.5$. Asymptotically, however, it turns out that he can essentially do just as well.

9.3.1 Nayak's No-Go Theorem

Theorem 1. *If any bit can be retrieved with probability $\geq 1-\epsilon$, then $n \geq m(1-H_2(\epsilon))$.*

To prove this, the following will be very useful.

Lemma 1. *Given states σ_0, σ_1 and a measurement $\{M_0, M_1\}$ which is good at distinguishing them, i.e., $Tr(M_b\sigma_b) \geq 1-\epsilon$ for each b, then*
$$S(\sigma) \geq \frac{S(\sigma_0)+S(\sigma_1)}{2} + 1 - H_2(\epsilon), \tag{9.45}$$
where
$$\sigma = \frac{\sigma_0+\sigma_1}{2}. \tag{9.46}$$

Proof. To prove the lemma, let us define the CQ state

$$\rho_{XQ} = \frac{|0\rangle\langle 0| \otimes \sigma_0 + |1\rangle\langle 1| \otimes \sigma_1}{2} \tag{9.47}$$

for which we know that

$$I(X:Q) = S(\sigma) - \left(\frac{S(\sigma_0) + S(\sigma_1)}{2}\right). \tag{9.48}$$

Considering the CQ channel/measure scenario

$$X - Q - B \tag{9.49}$$

given by

$$x \mapsto \sigma_x \mapsto b, \tag{9.50}$$

we know from the converse of the HSW theorem that the mutual information between X and B is at most $I(X:Q)_\rho$:

$$I(X:B) \leq I(X:Q)_\rho. \tag{9.51}$$

However, since $I(X:B)$ agree with probability at least $1-\epsilon$, we have $I(X:B) \geq 1 - H_2(\epsilon)$, from which the result follows. □

Finally, we can prove Nayak's No-Go Theorem.

Proof. We define the CQ state

$$\rho = \frac{1}{2^m} \sum_{x \in \{0,1\}^m} |x\rangle\langle x|^X \otimes \rho_x^Q. \tag{9.52}$$

Then, we know that there is a measurement that is good at distinguishing the cases of x_{k+1} being 0 or 1. The lemma then tells us that

$$S(Q|X_1 \dots X_k) \geq S(Q|X_1 \dots X_{k+1}) + 1 - H_2(\epsilon), \tag{9.53}$$

because $S(Q|X_1 \dots X_k)$ is the entropy of the state averaged over the values of X_{k+1}, and $S(Q|X_1 \dots X_{k+1})$ is the average of the entropies over the values of X_{k+1}. Iterating this gives

$$S(Q) \geq m(1 - H_2(\epsilon)), \tag{9.54}$$

and we conclude via the observation that $n \geq S(Q)$, since Q is an n-qubit register. □